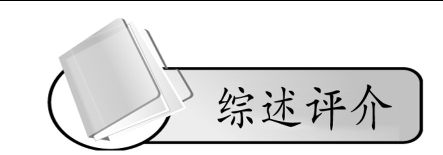


专家检索与专家排名研究评述*

叶光辉　夏立新

(华中师范大学信息管理学院　武汉 430079)

摘要: 【**目的**】对已有专家检索与专家排名方法进行评述，为后续研究提供理论基础。【**文献范围**】从 Web of Science (WOS)、CNKI 等数据库中分别以“专家检索(Expert Retrieval)”、“专家排名(Expert Ranking)”、“排名融合(Ranking Fusion)”等为检索词搜集获得相关文献 65 篇。【**方法**】针对专家检索覆盖面不足及专家特征计算量大两方面问题，从专家检索评测和排名融合两个角度梳理并评析现有的研究进展。【**结果**】融合关系属性是目前专家检索方法的主流，检索结果可信度研究主要依据用户满意度和文档可信度开展；专家排名采用友邻推荐模型、PageRank、D-S 理论、社交网络与复杂网络分析等实现排名及排名融合，融合结果总体优于基准排名。【**局限**】不同排名融合方法间的横向对比研究较少。【**结论**】相关研究可为构建信息融合视角下的专家会诊平台提供参考，具体体现在专家信息组织、专家遴选和专家意见融合环节。

关键词: 专家检索　排名融合　社交网络　关系属性　效果评测

分类号: G350

1 引　言

专家检索是将用户知识需求与潜在专家特征进行匹配的过程，常见的环节包括：用户知识需求表达、专家特征识别、知识需求与专家特征相似度计算以及候选专家结果输出等，但考虑到用户知识需求(尤其是应急知识需求)的时间敏感性和对检索结果的高要求，目前这一过程与用户快捷高效的要求并不匹配，造成这一现象的主要原因为检索系统专家类型覆盖面不足和专家特征识别计算工作量大，而这些问题在社交媒体时代尤其凸显。

专家类型覆盖面不足主要体现为：“小众专家”是社交网络环境下衍生的专家类型，其现实身份与虚拟身份可能存在极大反差，但他们直接参与到网络公共事件发生、发酵及衰退的过程中，对网络舆情的导控具有较强话语权，而现实公共事务咨询并没有将其视作对传统机构专家、学术专家的有益补充，从而导致专家类型覆盖面不足[1]，因此专家信息组织阶段要考虑“小众专家”作为专家库的组成部分，这样才能保证检索专家集合覆盖的完整性；专家特征识别计算工作量大主要体现为：在专家特征识别之前进行专家排名，可对特征识别对象进行有效的预选，即特征识别只识别综合排名较高的专家，而由于专家分面排名选用的指标多为显式的统计指标，因此可大大降低直接进行专家特征识别的计算量。在之前的研究中，笔者举例作出说明：假定专家检索获取专家集合容量为 n，与某一专家 e_i 关联的 x 类资源量为 q_i，对应专家组织需要提取的特征向量为 $\sum_{i=1}^{n} q_i$，获取这些特征向量后至少还需要$(2x-1)$次融合才能获取专家综合特征向量。而排名后再进行特征

通讯作者：叶光辉，OCRID：0000-0001-8111-5034，E-mail：3879-4081@163.com。

*本文系国家社会科学基金重大项目“基于多维度聚合的网络资源知识发现研究”(项目编号：13&ZD183)、中央高校基本科研业务费项目“面向应急决策的专家发现与意见融合研究”(项目编号：CCNU16A05044)和国家自然科学基金青年项目“多因素融合下的微博话题可信度评估模型及实证研究”(项目编号：71303179)的研究成果之一。

融合只需依据决策人员需求(假定为 r, $r<<n$), 选取排名比较靠前的若干位专家(大于等于 r)进行专家特征提取和融合, 需要提取的特征向量个数只有原来的 r/n[2]。

基于两方面问题分析, 笔者以 WOS、CNKI 等数据库为文献来源, 以专家检索、专家排名和排名融合为检索词, 尝试从专家检索方法及可信度评测和专家排名融合两个角度挖掘现有的研究方法、研究方案、研究工具等, 并有针对性地做出适时评析, 以为后期解决方案的衍生提供全面基础的参考。

2 专家检索与专家排名研究框架

专家检索与专家排名包括两种理解模式: 一种是先专家排名后检索模式, 另一种是先检索后专家排名模式, 如图 1 所示。

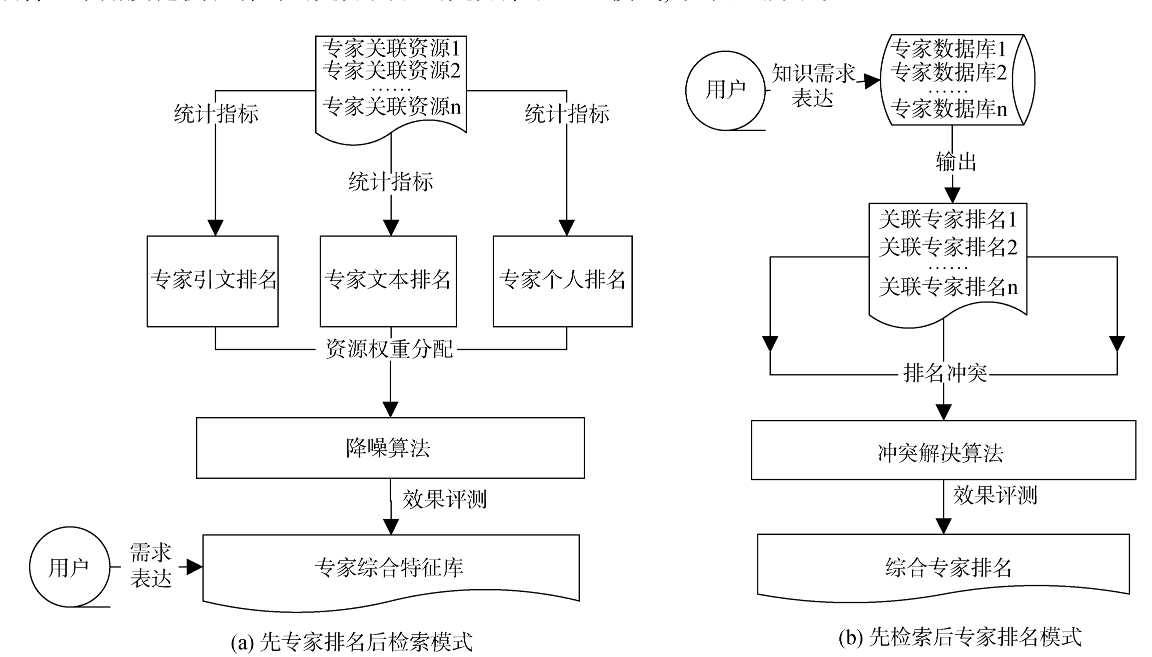


图 1 专家检索与专家排名的两种模式

图 1(a)和图 1(b)的差别是将专家排名融合放到信息组织阶段还是信息检索阶段, 这种差别也决定了最终要实现的目的。图 1(a)面向单一的专家检索系统, 期望在组织阶段融合多分面下专家排名值, 提升检索结果的准确性, 弱化信息噪声; 图 1(b)面向多专家检索或发现系统, 鉴于不同系统获取专家排名之间存在的冲突, 提出消解冲突获取综合排名的方法[2]。但无论哪一种模式, 专家检索与排名融合关联的研究主题基本一致: 专家检索可信度评测和专家排名融合。后续内容将不作服务目的界定, 单纯从研究主题角度评析现有的研究成果。

3 专家检索方法及可信度评测

为了更好地利用专家资源, 不少研究者都在积极构建科技咨询系统[3-4]、专家检索系统[5]、基于本体的专家定位系统[6]、FacFinder 专家搜索引擎[7]以及 SmallBlue 专家搜索软件[8-9]等。检索系统已成为定位和利用专家资源的重要途径, 目前国内三大学术数据库(CNKI、万方和维普)都提供学者或专家检索入口, 但数据量并不大。截至 2016 年 9 月 1 日, 万方共收录专家 12 120 位, 且专家信息组织偏重外部特征, 对内容特征的揭示不深。

3.1 专家检索方法

已有的专家检索研究往往与通用检索模型相结合, 按照用户实际需求进行适应性改造和创新。在之前的研究[10]中, 笔者指出专家特征信息包括两部分: 专家属性信息和专家关系信息。专家属性信息揭示专家自身特征, 专家关系信息源于专家在社交网络等公共平台上对音频、视频、文本、网页等资源的标注, 专家关系信息包括关系类型信息、标签信息、关系强度信息等[11]。依据检索算法是否考量专家关系属性, 可将现有专家检索方法分为: 融合

关系属性的专家检索方法和非融合关系属性的专家检索方法。

(1) 融合关系属性的专家检索方法

自 2005 年文本检索会议(TREC)发布专家搜寻的实验任务以来，相关研究者围绕相关实验数据集开展了一系列探索，产生了一批有影响力的成果，集中体现为: Fang 等[12-13]认识到在信息检索和机器学习领域，判别模型(Discriminative Mode)比现在应用较为广泛的生成模式(Generative Mode)表现效果更好，因此融合文档可信度及文档与专家关系强度，设计了一种基于判别模型的专家检索方法，并通过 TREC 提供的数据集，证明该方法的有效性和健壮性; MacDonald 等[14]设计了基于文档可信度的专家搜索方法: 给定检索提问，检索系统首先考虑文档与检索提问的相关程度，对同一专家的不同文档给予可信度评估，再依托文档可信度差异，辅以查询扩展等，获取专家最终排名，并通过 TREC2005 和 TREC2006 相关数据集，证明该方法具有较好的性能表现; Zhai 等[15]带领 UIUC 团队参与 TREC 测试，并设计了一种基于语言模型(Language Model)的专家发现方法，而 Zhu 等[16]认为专家发现与其关联文档特征的揭示完备性有很强关系，而已有的语言模型没有充分考虑到这些特征，故提出了一种整合多重文档特征的新型语言模型，并以 W3C 和 CSIRO(TREC)为数据集，证明该方法的平均准确率(Mean Average Precision, MAP)比原语言模型方法要高。

除 TREC 外，其他研究者也做出了很多典型工作: Uddin 等[17-18]设计了一种专家搜索方法，将本体构建与科研社区网络方法相结合，利用本体去描述科研社区网络中节点的属性及关系，以专家相关成果及影响得分为排名依据; Zhou 等[19]发现在科研问答社区专家发现过程中，采用链接分析方法易忽略专家兴趣、声望等，提出基于主题概率模型的专家发现方法，并通过扩展 PageRank 算法证实该方法可获得优良的检索效果; 王雪芬等[20]认识到当前专家组织与检索过程中存在的问题，结合社会网络，融合专家各类特征信息，提出较为新颖的专家检索技术方案; Farhadi 等[21]从社交问答平台中发现，用户有时需求的专家数量不止一人，而是一个专家团队，为此设计基于共现[22]和聚类的社交网络专家发现的框架: TeamFinder，并以 DBLP 为数据集，借助作者共现说明了框架的有用性和适用范围; Sun 等[23]鉴于基于图检索的模型没有充分考虑一些重要的上下文信息，融合专家质量及专家间关系，提出一种新颖的科研社区专家发现方法，并通过实验证实所提方法相较基准方法具有更好的推荐效果; 赵红斌等[24]就专家识别过程中出现的一词多义、异词同义、专有名词等自然语言分析问题，构建专家专长词表来辅助专家专长识别，有效地提升了专家检索的准确率，并以武汉大学教师群体为例，构建相关原型系统—武汉大学专家检索系统 WHU-ES。

(2) 非融合关系属性的专家检索方法

相比融合关系属性的专家检索方法的多样性，非融合关系属性的专家检索方法则较少: Van Gysel 等[25]完全单一借助专家文本信息，引入一种基于无监督判别模型的专家检索方法，对比基于向量空间和概率生成模型的专家检索结果，发现该方法生成专家的排名效果明显优于后两者; Yang 等[26]基于普遍的共识: 专家所具有的隐性知识不易系统化和结构化，搜寻相关主题专家比搜寻知识管理系统储存的知识(Knowledge Management System, KMS)成本要低，采纳基于改进型 Fuzzy Abstraction Hierarchy 框架的智能专家搜寻方法，并开发了相关的原型系统; Stankovic 等[27]鉴于目前关联数据网络(Linked Data Web)使得用户社交行为可追踪和现有专家检索使用单一语料集的现实情况，提出利用关联数据计量分析方法，选择正确的跟踪类型和专业知识假说，获取准确率和召回率较高的关联专家; 方锴[28]和武浩等[29]针对专家检索在科学研究、企业管理等场合的重要应用价值，对专家检索的研究目标、内容和方法进行梳理，重点归纳了专家描述、查询主题与专家关系建模、链接分析、查询扩展和专家证据识别等主题，并预估专家检索未来将切入不同应用场景，为科研团队发现、企业知识管理等提供切实帮助; 陈霄咚等[30]认识到目前专家寻找的方法各有优缺点，如语言模型、主题模型等，提出一种可有效结合已有专家寻找模型的融合框架，并通过实验证明该框架可有效提高专家寻找的精确度与鲁棒性。

综上分析可知，目前专家检索方法研究倾向于融合关系属性的专家检索方法，这一方面源于社交网络媒体的飞速发展，另一方面也在于专家关联资源的日趋多源。综合多渠道信息资源提取的专家特征，在实

现专家信息完备的同时也造成了较大的信息噪声和冲突[31]，因此专家检索不仅要在源头上强化专家信息组织方法和技术，还要针对专家关联资源设计可信度测评机制；同时由于专家检索存在多种 ExpertRank[32]算法，但这些算法只是将检索用户作为接收检索结果的客体，没有为检索用户设置参与算法执行的变量，用户自主检索功能不强，更无从谈及个性化专家检索。因此，专家检索研究不应完全将重心集中于通用检索模型的改进工作，随着用户体验在检索过程中越来越受到关注，专家检索输出应更多考虑用户评分及关联资源的可信度。

3.2　专家检索可信度评测

基于已有专家检索系统获取的检索结果的可信度，其大小可作为结果相关度判断的标准之一，为用户接受或拒绝检索信息提供参考。前期专家信息有效的组织才能保证后期专家遴选过程的科学高效，如果说专家组织保证了遴选专家即信源的可信度[33-34]，则在专家排名[31]、群体决策[35-36]以及同行评议[37-38]等咨询活动中，为获取高效统一的决策，围绕专家意见或观点所应用的方法、模型、算法、系统等可划定为是对信源内容所做的可信度评估。目前可信度评估研究的对象主要为信息资源，特别是网络信息资源，这主要源于大量冗余、虚假和模糊的信息充斥于网络环境中，而专家资源作为信息资源的重要组成部分，其采用的可信度评估方式主要包括两类：基于用户满意度的可信度评估和基于文档的可信度评估。

(1) 基于用户满意度的可信度评估

基于用户满意度的可信度评估主要是从用户角度对专家检索效果进行评估，可以是系统外的评估，如通过实验法、问卷法、访谈法等获取用户对专家检索系统的主观印象，也可以是系统内评估，如用户根据自身检索需求，自主设定检索算法中参数，从而达到改变检索进程和检索结果的目的，而检索进程的改变可以始于信息组织层面，也可以施行于信息检索层面。李纲等[39]认为用户主导下系统内信息组织的改变可称为后端可信度评估，系统内信息检索层面的改变可被称为前端可信度评估，后端可信度评测通过最佳专家特征向量目长的求解[40]来降低检索噪声，前端可信度评测将用户相关性反馈作为检索路径选择的必要参照；Kim 等[41]以大学生及专家为访谈对象，对其在在线健康信息搜寻中的行为进行分析，证明良好的网站外观和导航设计有助于用户获得较好的检索体验；Wu 等[42-43]在原有企业搜索服务(包括：服务提供者、检索系统和服务使用者)中引入中间环节：调和企业搜寻，它利用领域专家经验知识对搜索结果进行可信度分析，最后再推送给用户，并证明改进后的服务有助于服务开发者获得长久的投资回报和提供更为准确的搜索结果；Liebregts 等[44]构建了一个大学专家搜索系统，以科研人员、学生和机构外访问者为检索用户，通过提问评测证实该系统具有不错的系统效率、检索效率和用户满意度；Jefferson 等[45]提出一种利用信息检索技术和领域专家知识来改善用户检索满意度的方法，并依托该方法设计了带有智能前端的原型系统，评测证实该系统检索效率要高于传统的目录系统。

(2) 基于文档的可信度评估

基于文档的可信度评估较基于用户满意度的可信度评估更为客观，主要研究体现为：MacDonald 等[46-47]指出文档排名与专家排名具有很强的关系，也是专家检索效率提升的重要因素，并通过一次专家检索实验失败的原因分析说明文档可信度评估对专家检索的影响；Kim 等[48]发现目前多数自动排名系统使用引用关系描述网络文档的相对重要程度，但引用关系无法反映用户评论的内容，即只是单纯的计量，不涉及语义层面，为此设计一种基于特定网页集合自动生成专家团队的方法，依据用户评论可信度和关联网页集合计算每位专家的影响力权值；Noll 等[49-50]指出在协同标注系统中发现指定主题的专家，决定于两个因素：一是用户发布相关的质量较高的资源，而资源质量依赖用户的标注行为，另一个是该用户领先其他用户关注或识别到相关资源，基于此设计基于图模型的专家发现算法：SPEAR(Spamming-Resistant Expertise Analysis and Ranking)，并以 Delicious(https:// del.icio.us/)网页、用户和书签为数据源，证实该算法在大众分类网站中具有良好效果。

由上述分析可知，对应 3.1 节专家检索方法分类，专家检索可信度评测方法包括两类：基于用户满意度的可信度评估和基于文档的可信度评估。前者不再将重点集中于通用检索模型的改进工作，而是更多考虑用户评分，甚至为检索用户设置参与算法执行的变量，提升用户自主检索功能；后者要在源头上强化专家信

息组织方法和技术, 针对专家关联资源设计可信度测评机制。

4 专家排名融合

专家排名是对专家综合影响力的量化, 但利用不同资源(如学术网络、Web 资源、社会网络资源)获取的专家排名不一定一致, 甚至可能存在冲突, 为此需要将基于不同统计指标获取的专家分面排名进行融合, 从而获取专家的综合排名。同时, 利用简单的统计指标先对关联主题的专家进行排名, 还可以起到专家预选或筛选的目的, 大大减轻专家特征融合的任务量。排名所用的统计指标可来自个人信息, 如发文量、最初发表时间等, 也可来自引文信息, 如 H 指数、被引频次等, 但无论哪一种形式的排名, 为实现排名融合都需要进行标准化处理。其实多数专家检索系统, 排名已被设定在检索算法中。目前, 该主题研究主要围绕专家排名及排名融合方法展开。

4.1 专家排名方法

专家排名方法研究主要体现为: Wu 等[51]描述了一种基于几何框架的数据融合方法, 将来自不同检索系统的数据表示成多维空间的一个点, 通过欧几里得距离(Euclidean Metric)度量检索结果的相似度, 并与已有融合方法作比较, 证实该融合方法更为准确和高效; Uddin 等[17]指出专家发现是社交网络研究的重要主题, 但很少有研究从语义层面来对此进行探讨, 为此提出基于本体的专家发现方法, 通过专家对主题相关文档的贡献和专家间关系计算专家得分, 据此生成最优专家排名; Wang 等[52]指出企业微博商务管理具有很强的重要性, 尤其是在社交网络服务日益发展的今天, 因此结合 PageRank 算法和专家标注的标签, 设计企业专家排名的方法; Jin 等[53]指出随着社交网络日益流行, 寻找解决问题的专家可能比直接搜索答案更易实现, 为此将用户间的中心度计算、提问与用户之间的相关性加总在一起, 从而获取用户需求的专家排名结果; Moreira 等[54]鉴于目前学术数字图书馆仍然缺乏一种有效的融合多种显式特征的专家发现算法, 从专家关联资源的上下文、专家个人信息、科研社区中基于引链关系形成的图结构出发, 构建了一种规范化的专家融合及排名方法, 并以计算机科学领域为例证实了该方法的性能; Wu 等[55]认为已有专家发现方法使用面不广, 只能满足特定的学术需求, 在这些方法中作者间的合著关系和引用关系是评估专家的重要考量内容, 为此提出一种基于灵活数据模式且主题关联的共 PageRank 算法, 算法将引文网络和社会网络上专家与主题的偏好程度作为专家排名计算的依据; 詹镇江[56]认为传统的专家搜索更多向用户推送相关主题专家的专业技能, 而忽略专家与用户之间的关系, 为此提出融入用户社交网络的社会化搜索方式, 通过用户与主题的相关程度及用户在网络中的影响力以计算专家排名, 并证实该方法在友邻推荐模型(Friend Recommendation Model, FRM)中具有较好的应用效果; 张波[57]充分考虑用户发布过的微博内容和用户之间的拓扑关系, 通过引入主题模型来拓展 PageRank 算法, 计算用户在各个主题下的社交影响力, 最终将主题社交影响力和内容相关度相结合来给出较为准确的专家定位排名, 并利用新浪微博 API 抓取的实验数据集进行实验, 证实了该方法的有效性。

4.2 排名融合方法

除了资源多样性为排名融合提供了研究空间, 既有的采用单一指标的排名还存在最终排名结果可信度不强、排名不一致、排名冲突等问题。为解决上述问题, 图书情报领域衍生出学科内的一套方法, 在部分研究中, 为减少数据的稀疏性, 需要抽取高频作者以构建关系矩阵, 而高频作者抽取可基于阈值设计的主观方法, 如设定一个抽取的最低频次, 也可根据数据的分布规律, 如符合幂律分布, 则可使用普赖斯公式来提取。无论哪一种方法, 目前都不够完善。设置最低阈值仅仅是为了满足数据分析的需求, 缺乏训练集的支撑, 而且无法解决末端数据一致的问题, 同时这也是排名指标要多源的一个重要原因, 而客观提取方式又对数据的分布情况具有严格的要求。

为此, 专家排名融合研究积极借鉴和改造不同学科领域(如遥感测绘、人脸识别、专家系统等)的融合方法, 如 Santos 等[58]通过对搜索系统 WSEs 的专家搜索模块进行调研后指出: 用于专家检索的文档也要给予一定可信度, 融合已有的专家排名指标, 并利用 TREC 数据生成 7 种不同专家排名, 证明了融合文档可信度的专家排名在企业内网检索过程中获得了优良的效果; MacDonald 等[59]认为专家排名可以采用投票模型来解释, 尝试利用 11 种数据融合技术完成专家排

名, 并使用 TREC 企业相关数据集, 证明了该专家排名融合方法可有效改善候选专家的排名, 具有相对稳定的检索性能; Mourão 等[60]指出医学领域多源异构信息的不断涌现给相关检索带来很大挑战, 为此提出可支持多维数据融合、词表荐词的医学信息检索系统构建模式, 实验表明新的融合算法可有效改善已有的排名融合结果, 搜索效果优于其他医学检索系统, 如 2013 年的 ImageCLEFMedical; Yang 等[61]认为随着信度函数理论(Belief Functions Theory)识别框架中元素数量的不断增加, 证据合并过程中造成的计算成本也会不断增加, 为此尝试从元素数量和 Mass 值分配两个角度出发降低计算成本, 并借助相关实验及分析证实所提基本概率分配(The Basic Probability Assignment)约值法的性能表现; Nuray 等[62]提出一种新的检索排名方法: 通过多源数据融合算法将来自不同检索系统的相关文档合并, 然后选取 Top N 位的文档作为伪相关反馈的实验文档, 通过它们去评估各个检索系统并获取系统排名, 利用 TREC 数据集进行实证分析, 发现分析结果与人工评估结果呈现较高的一致性; Franceschini 等[63-64]指出在多准则决策与社会选择理论等研究领域, 常常由于不同代理的偏好而造成备选方案的顺序不一致, 为此提出了一种新型的融合算法: Ordered Paired-Comparisons Algorithm(OPCA), 算法包括三个步骤: 首先将产生的多代理排名转化为成对比较的集合, 然后基于一定的优先级原则融合这些集合, 最后获取代理人重要性的排名, 生成一个统一的次序, 并通过实例证实了该方法的融合效果; Zhao 等[65]为解决社交网络平台中个性化标签推荐问题, 提出一种新型的服务于多源信息的排名框架 GRoMO (Graph-based Ranking of Multi-type Interrelated Objects), 该框架对标签标记资源与用户标记的历史记录进行分析, 通过特定的推荐算法向用户推送排名 Top N 的标签, 并爬取知名社交平台 Delicious 数据进行实证分析, 证实该方法效果优于基准方法, 且随着数据集的不断扩展, 也可适用于资源推荐。

4.3 排名融合研究评析

专家排名融合包括专家排名方法及排名融合两部分。由 4.1 节和 4.2 节可知, 现有专家排名方法具有 4 方面特征:

(1) 多数专家排名方法将专家属性特征信息与关系特征信息相结合, 采用多种指标作为设置排名的参考;

(2) 专家排名值计算方法则不一而足, 包括加权平均、友邻推荐模型、投票模型、PageRank、D-S 理论、社交网络与复杂网络分析等;

(3) 排名对象包括企业专家、学术专家和社交网络专家, 其中社交网络专家研究比较广泛, 源于围绕这些专家节点及其社交行为易于构建社会语义网络, 便于通过网络科学理论和指标进行方法设计;

(4) 排名效果在解决排名冲突的前提之下, 都比基于单一指标的排名方法效果要好, 但由于排名数据集的差异, 改进后不同排名方法的横向比较还未进行深入的探讨与分析。

5 结　语

笔者提出已有专家检索中存在的两类问题: 一是专家覆盖面不足; 二是专家特征计算量大, 并对专家检索方法、可信度评测与排名融合相关研究成果进行系统梳理和评析。专家检索评测主要围绕专家检索方法与检索可信度评测方法展开。现有的专家检索方法往往在通用检索模型基础上加以改造, 融合关系属性的专家检索方法是该主题研究的主流。与之相对照, 专家检索可信度评测方法也包括两类: 基于用户满意度的可信度评估和基于文档的可信度评估, 前者注重用户评分, 后者看重资源质量。专家排名融合包括专家排名方法及排名融合两部分。专家排名方法以关系特征为主, 属性特征为辅, 采用加权平均、友邻推荐模型、投票模型、PageRank、D-S 理论、社交网络与复杂网络分析等方法实现排名及排名融合, 实证结果显示改造后的排名结果总体优于基于单一指标的排名结果, 但不同排名方法间的横向对比研究则较少。

结合各个研究环节的成果及笔者正在开展的研究工作, 与之对应的解决方案也可大体描述为: 融合学术资源、Web 资源、社交网络资源共同作为传统专家特征识别的关联资源, 设计加权平均等多源信息融合方法以解决专家特征融合过程出现的噪声和冲突[10], 并在处理网络舆情事件中, 引入"小众专家"构建专家库; 结合专家检索方法与排名融合算法, 在专家检索之前通过专家分面排名的融合既可提升专家信息组织的准确性和完备性[2], 也可通过有效预选大大降低特

征计算的工作量，最终检索结果可通过基于文档或用户满意度的个性化评测算法实现二次反馈，从而改善专家检索的可信度[39]。

结合已有研究现状及后续研究思路，笔者认为专家检索与排名研究将呈现以下发展趋势:

随着专家关联资源的日趋多样以及移动社交问答等平台的涌现，现有专家定义的外延将被进一步拓展，互联网环境下的"小众专家"将愈发活跃，传统机构专家、技术专家与"小众专家"的交集将越来越大。

资源多样势必带动多源信息融合，专家特征信息融合过程要重点解决两类问题: 信息噪声和信息冲突，围绕它们进行算法、模型等改进或创新工作将是专家检索与排名研究未来发展的主要趋势，进而带动专家检索个性化、语义化和精准化服务水平的提升。

参考文献:

[1] 李纲，叶光辉，张岩. "小众专家"特征识别研究——基于MetaFilter 的实证分析[J]. 现代图书情报技术，2015(6): 71-77. (Li Gang, Ye Guanghui, Zhang Yan. Feature Recognition of Niche Expert——Empirical Analysis Based on MetaFilter Dataset[J]. New Technology of Library and Information Service, 2015(6): 71-77.)

[2] 叶光辉，李纲，武川. 应急参考咨询团队构建模式研究[J]. 情报学报, 2015, 34(7): 734-742. (Ye Guanghui, Li Gang, Wu Chuan. Research on Construction Mode of Emergency Experts Team for Virtual Reference[J]. Journal of the China Society for Scientific and Technical Information, 2015, 34(7): 734-742.)

[3] 王曰芬，王雪芬，杨小晓. 基于社会网络的科技咨询专家库的构建方案和流程设计[J]. 情报学报，2012, 31(2): 116-125. (Wang Yuefen, Wang Xuefen, Yang Xiaoxiao. Research on Construction Schema and Program Design of Social Network-based Expert Database in Scientific and Technical Consulting[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(2): 116-125.)

[4] 王曰芬，王雪芬，颜端武. 基于社会网络的科技咨询专家库的原型系统设计与实现[J]. 情报学报，2012, 31(3): 250-258. (Wang Yuefen, Wang Xuefen, Yan Duanwu. Prototype System Design and Implement of the Database About Science and Technical Consulting Experts Based on Social Network[J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(3): 250-258.)

[5] 陆伟，韩曙光. 组织专家的检索系统设计与实现[J]. 情报学报, 2008, 27(5): 657-663. (Lu Wei, Han Shuguang. Design and Implementation of Organization Expert Search System[J]. Journal of the China Society for Scientific and Technical Information, 2008, 27(5): 657-663.)

[6] 刘萍，叶燕. 基于本体的高校专家定位系统研究[J]. 情报学报，2010, 29(5): 813-819. (Liu Ping, Ye Yan. An Ontology-based Experts Locator System within Academia[J]. Journal of the China Society for Scientific and Technical Information, 2010, 29(5): 813-819.)

[7] Fang Y, Si L, Mathur A. FacFinder: Search for Expertise in Academic Institutions[R]. West Lafayette: Purdue University, 2008.

[8] Ehrlich K, Lin C Y, Griffiths-Fisher V. Searching for Experts in the Enterprise: Combining Text and Social Network Analysis[C] // Proceedings of the 2007 International ACM Conference on Supporting Group Work. New York: ACM, 2007: 117-126.

[9] Lin C Y, Ehrlich K, Griffiths-Fisher V, et al. SmallBlue: People Mining for Expertise Search[J]. IEEE MultiMedia, 2008, 15(1): 78-84.

[10] 李纲，叶光辉. 多源专家特征信息融合研究[J]. 现代图书情报技术, 2014(4): 27-33. (Li Gang, Ye Guanghui. Research on Information Fusion for Multiple-Sensor Expert Features [J]. New Technology of Library and Information Service, 2014(4): 27-33.)

[11] 彭红彬，王军. 虚拟社区中知识交流的特点分析——基于CSDN 技术论坛的实证研究[J]. 现代图书情报技术，2009(4): 44-49. (Peng Hongbin, Wang Jun. Topology of the Knowledge Communication Network in Virtual Communities Based on CSDN [J]. New Technology of Library and Information Service, 2009(4): 44-49.)

[12] Fang Y, Si L, Mathur A P. Discriminative Probabilistic Models for Expert Search in Heterogeneous Information Sources [J]. Information Retrieval Journal, 2011, 14(2): 158-177.

[13] Fang Y, Si L, Mathur A P. Discriminative Models of Integrating Document Evidence and Document-Candidate Associations for Expert Search[C]//Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2010: 683-690.

[14] MacDonald C, Ounis I. Using Relevance Feedback in Expert Search[C]// Proceedings of the European Conference on Information Retrieval. Berlin: Springer, 2007: 431-443.

[15] Fang H, Zhou L, Zhai C X. Language Models for Expert Finding--UIUC TREC 2006 Enterprise Track Experiments

[C]//Proceedings of 15th Text Retrieval Conference(TREC 2006), Gaithersburg, Maryland, USA. 2006.

[16] Zhu J, Huang X, Song D, et al. Integrating Multiple Document Features in Language Models for Expert Finding [J]. Knowledge and Information Systems, 2010, 23(1): 29-54.

[17] Uddin M N, Duong T H, Oh K, et al.Expert Search and Rank with Social Network: An Ontology-Based Approach [J]. International Journal of Software Engineering and Knowledge Engineering, 2013, 23(1): 31-50.

[18] Uddin M N, Duong T H, Oh K, et al. An Ontology Based Model for Experts Search and Ranking [A]// Lecture Notes in Computer Science [M]. New York: ACM, 2011, 6592(6): 150-160.

[19] Zhou G, Zhao J, He T, et al. An Empirical Study of Topic-Sensitive Probabilistic Model for Expert Finding in Question Answer Communities [J]. Knowledge-Based Systems, 2014, 66(9): 136-145.

[20] 王雪芬, 王曰芬. 专家库中的专家检索技术研究[J]. 情报理论与实践, 2011, 34(2): 96-99. (Wang Xuefen, Wang Yuefen.Research on Expert Retrieval Technical in Expert Database [J]. Information Studies: Theory & Application, 2011, 34(2): 96-99.)

[21] Farhadi F, Hoseini E, Hashemi S, et al. TeamFinder: A Co-clustering Based Framework for Finding an Effective Team of Experts in Social Networks[C]. In: Vreeken J, Ling C, Zaki M J, et al. Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops. Washington: IEEE Computer Society, 2012: 107-114.

[22] Guan Z, Miao G, McLoughlin R, et al. Co-Occurrence-Based Diffusion for Expert Search on the Web [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(5): 1001-1014.

[23] Sun J, Xu W, Ma J, et al. Leverage RAF to find Domain Experts on Research Social Network Services: A Big Data Analytics Methodology with MapReduce Framework [J]. International Journal of Production Economics, 2015, 165: 185-193.

[24] 赵红斌, 陆伟. 专家研究领域自动识别研究[J]. 现代图书情报技术, 2010(2): 63-67. (Zhao Hongbin, Lu Wei. The Study of Expert Research Field Automatic Recognition [J]. New Technology of Library and Information Service, 2010(2): 63-67.)

[25] Van Gysel C, De Rijke M, Worring M. Unsupervised, Efficient and Semantic Expertise Retrieval[C]// Proceedings of the 25th International Conference on World Wide Web. New York: ACM, 2016: 1069-1079.

[26] Yang K, Huh S. Intelligent Search for Experts Using Fuzzy Abstraction Hierarchy in Knowledge Management Systems [J]. Journal of Database Management, 2007, 18(3): 47-68.

[27] Stankovic M, Jovanovic J, Laublet P. Linked Data Metrics for Flexible Expert Search on the Open Web[C]//Proceedings of the Extended Semantic Web Conference. 2011, 6643: 108-123.

[28] 方锴. 专家检索研究综述[J]. 科协论坛, 2013(2): 102-104. (Fang Kai. A Review on Expert Retrieval [J]. Science & Technology Association Forum, 2013(2): 102-104.)

[29] 武浩, 王美姣, 冯佳明, 等. 专家检索研究进展[J]. 计算机应用研究, 2010, 27(10): 3633-3638. (Wu Hao, Wang Meijiao, Feng Jiaming, et al. Research Advance of Expert Finding[J]. Application Research of Computers, 2010, 27(10): 3633-3638.)

[30] 陈霄咚, 丁浩. 专家寻找模型融合框架研究[J]. 计算机应用与软件, 2014, 31(12): 74-79. (Chen Xiaodong, Ding Hao. On Fusion Framework for Expert Finding Models[J]. Computer Applications and Software, 2014, 31(12): 74-79.)

[31] Moreira C, Wichert A. Finding Academic Experts on a MultiSensor Approach Using Shannon's Entropy [J]. Expert Systems with Applications, 2013, 40(14): 5740-5754.

[32] Wang G A, Jiao J, Abrahams A S, et al. ExpertRank: A Topic-aware Expert Finding Algorithm for Online Knowledge Communities[J]. Decision Support Systems, 2013, 54(3): 1442-1451.

[33] 王平, 程齐凯. 网络信息可信度评估的研究现状及述评[J]. 信息资源管理学报, 2013, 3(1): 46-52. (Wang Ping, Cheng Qikai. Review and Progress in Research on Credibility Evaluation of Information on the Web [J]. Journal of Information Resources Management, 2013, 3(1): 46-52.)

[34] 邓发云. 基于用户需求的信息可信度研究[D]. 成都: 西南交通大学, 2006. (Deng Fayun. Information Credibility on User Demand[D]. Chengdu: Southwest Jiaotong University, 2006.)

[35] 梁樑, 熊立, 王国华. 一种群决策中确定专家判断可信度的改进方法[J]. 系统工程, 2004, 22(6): 91-94. (Liang Liang, Xiong Li, Wang Guohua. A New Method of Determining the Reliability of Decision-Makers in Group Decision [J]. Systems Engneering, 2004, 22(6): 91-94.)

[36] Zhao H X, Yang J P, Wang J. Evaluation Model of Credibility of E-Commerce Website Using Fuzzy Multi-Attribute Group Decision Making: Based on Fuzzy Structured Element[C]// Proceedings of the Fuzzy Information and Engineering 2010.

Berlin & Heidelberg: Springer, 2010: 417-424.
[37] 徐林生, 王执铨, 戴跃伟. 评审专家可信度评价模型及应用[J]. 南京理工大学学报: 自然科学版, 2010, 34(1): 30-34. (Xu Linsheng, Wang Zhiquan, Dai Yuewei. Expert's Credibility Model and Evaluation Method[J]. Journal of Nanjing University of Science and Technology: Natural Science, 2010, 34(1): 30-34.)
[38] 李振清, 刘建毅, 王枞, 等. 同行评议专家遴选系统研究与实现[J]. 现代图书情报技术, 2012(5): 81-86. (Li Zhenqing, Liu Jianyi, Wang Cong, et al. Research and Implementation of Peer-Review Experts Selection System [J]. New Technology of Library and Information Service, 2012(5): 81-86.)
[39] 李纲, 叶光辉.用户主导下的专家检索可信度评测机制研究[J]. 现代图书情报技术, 2014(7/8): 107-113. (Li Gang, Ye Guanghui. Research on Credibility Evaluation Mechanism of Experts Retrieval Under User's Control [J]. New Technology of Library and Information Service, 2014(7/8): 107-113.)
[40] Börner K, Dall'Asta L, Ke W, et al. Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-authorship Teams[J]. Complexity, 2005, 10(4): 57-67.
[41] Kim H, Park S Y, Bozeman I. Online Health Information Search and Evaluation: Observations and Semi-Structured Interviews with College Students and Maternal Health Experts[J]. Health Information & Libraries Journal, 2011, 28(3): 188-199.
[42] Wu M, Thom J A, Turpin A, et al. Cost and Benefit Analysis of Mediated Enterprise Search[C]//Proceedings of the 2009 ACM/IEEE Joint Conference on Digital Libraries.New York: ACM, 2009: 267-276.
[43] Wu M, Turpin A, Thom J A, et al. Cost and Benefit Estimation of Experts' Mediation in an Enterprise Search [J]. Journal of the Association for Information Science & Technology, 2014, 65(1): 146-163.
[44] Liebregts R, Bogers T. Design and Evaluation of a University-Wide Expert Search Engine [C]// Proceedings of the European Conference on Information Retrieval.Springer, 2009: 587-594.
[45] Jefferson T I, Nagy T J. A Domain-Driven Approach to Improving Search Effectiveness in Traditional Online Catalogs [J]. Information & Management, 2002, 39(7): 559-570.
[46] MacDonald C, Ounis I. The Influence of the Document Ranking in Expert Search [J]. Information Processing & Management, 2011, 47(3): 376-390.
[47] MacDonald C, Ounis I. On Perfect Document Rankings for Expert Search[C]//Proceedings of the 32nd International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM, 2009: 740-741.
[48] Kim S W, Chung C W. Ranking Web Documents with Dynamic Evaluation by Expert Groups[C]//Proceedings of the International Conference on Advanced Information Systems Engineering. Berlin: Springer, 2003: 437-448.
[49] Noll M G, Au Yeung C M, Gibbins N, et al. Telling Experts from Spammers: Expertise Ranking in Folksonomies[C]// Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2009: 612-619.
[50] Noll M G, Meinel C. Web Search Personalization via Social Bookmarking and Tagging[C]//Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference.2007: 367-380.
[51] Wu S, Crestani F. A Geometric Framework for Data Fusion in Information Retrieval [J]. Information Systems, 2012, 50: 20-35.
[52] Wang Y, Zhang L J, Ning K, et al. Experts Ranking on the Enterprise Microblogging Based on the PageRank Algorithm [C]//Proceedings of the 2012 IEEE Asia-Pacific Services Computing Conference. Washington: IEEE Computer Society, 2012: 345-349.
[53] Jin L, Yoon J Y, Kim Y H, et al. Based on Analyzing Closeness and Authority for Ranking Expert in Social Network [C]//Proceedings of the 7th International Conference on Advanced Intelligent Computing Theories and Applications: with Aspects of Artificial Intelligence.Berlin: Springer, 2011: 277-283.
[54] Moreira C, Calado P, Martins B. Learning to Rank for Expert Search in Digital Libraries of Academic Publications[C]// Proceedings of the Portuguese Conference on Artificial Intelligence. Berlin: Springer, 2011: 431-445.
[55] Wu H, Li H, Zhang X, et al. Topic-Sensitive Link-Ranking Approach for Academic Expert Recruiting[C]//Proceedings of the 2008 International Multi-Symposiums on Computer and Computational Sciences. Washington: IEEE Computer Society, 2008: 150-157.
[56] 詹镇江. 基于专家与用户关系的社会化专家搜索[D]. 上海: 上海交通大学, 2012. (Zhan Zhenjiang. Socialized Expert Finding Based on the Relationship Between Experts and the Users [D]. Shanghai: Shanghai Jiaotong University, 2012.)

[57] 张波. 基于 LDA 的微博服务专家定位方法研究[D]. 上海: 华东师范大学, 2013. (Zhang Bo.The Research on Expert Location in Microblogging Based on LDA [D]. Shanghai: East China Normal University, 2013.)

[58] Santos R L T, MacDonald C, Ounis I. Mimicking Web Search Engines for Expert Search [J]. Information Processing & Management, 2011, 47(4): 467-481.

[59] MacDonald C, Ounis I.Voting Techniques for Expert Search [J]. Knowledge & Information Systems, 2008, 16(3): 259-280.

[60] Mourão A, Martins F, Magalhães J. Multimodal Medical Information Retrieval with Unsupervised Rank Fusion [J]. Computerized Medical Imaging & Graphics, 2014, 39: 35-45.

[61] Yang Y, Han D, Han C, et al. A Novel Approximation of Basic Probability Assignment Based on Rank-Level Fusion [J]. Chinese Journal of Aeronautics, 2013, 26(4): 993-999.

[62] Nuray R, Can F. Automatic Ranking of Information Retrieval Systems Using Data Fusion [J]. Information Processing & Management, 2006, 42(3): 595-614.

[63] Franceschini F, Maisano D, Mastrogiacomo L. A Paired-Comparison Approach for Fusing Preference Orderings from Rank-Ordered Agents [J]. Information Fusion, 2015, 26: 84-95.

[64] Franceschini F, Maisano D, Mastrogiacomo L. A Novel Algorithm for Fusing Preference Orderings by Rank-ordered Agents [J]. Fuzzy Sets & Systems, 2015, 266: 84-100.

[65] Zhao W, Guan Z, Liu Z. Ranking on Heterogeneous Manifolds for Tag Recommendation in Social Tagging Services [J]. Neurocomputing, 2015, 148: 521-534.

作者贡献声明:

叶光辉: 思路及命题制订, 资料收集, 论文撰写及文字校对;
夏立新: 论文终稿修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

收稿日期: 2016-09-12
收修改稿日期: 2016-12-13

Review of Expert Retrieval and Expert Ranking Studies

Ye Guanghui Xia Lixin
(School of Information Management, Central China Normal University, Wuhan 430079, China)

Abstract: [**Objective**] This paper reviews the expert retrieval and expert ranking literature to provide theoretical foundations for future studies. [**Coverage**] 65 papers were retrieved from the Web of Science (WOS), CNKI and other databases using the keywords of “expert retrieval”, “expert ranking”, and “ranking fusion”. [**Methods**] We analyzed research evaluating expert retrieval and fusion rankings, aiming to solve the issues of insufficiency of expert coverage and heavy computation of expert features. [**Results**] We found that most expert retrieval system adopted the relationship attribute fusion method, and the credibility of search results was decided by the users’ satisfaction and quality of the retrieved documents. Expert ranking was established by FRM, PageRank, D-S theory, social network and complex network analysis. Empirical research showed that the fusion ranking results were generally better than the baseline ones. [**Limitations**] More comparison of research among different ranking methods was needed. [**Conclusions**] Related studies help us building expert consulting platform from the perspective of expert information organization, expert selection and expert opinion fusion.

Keywords: Expert Retrieval Ranking Fusion Social Network Relationship Attribute Effect Evaluation